

THE BASQUE LANGUAGE
IN THE DIGITAL AGE

EUSKARA
ARO
DIGITALEAN

Inmaculada Hernáez
Eva Navas
Igor Odriozola
Kepa Sarasola
Arantza Diaz de Ilarraza
Igor Leturia
Araceli Diaz de Lezana
Beñat Oihartzabal
Jasone Salaberria



White Paper Series

Liburu Zurien Bilduma

THE BASQUE LANGUAGE IN THE DIGITAL AGE

EUSKARA ARO DIGITALEAN

Inmaculada Hernáez [1]

Eva Navas [1]

Igor Odriozola [1]

Kepa Sarasola [1]

Arantza Diaz de Ilarraza [1]

Igor Leturia [2]

Araceli Diaz de Lezana [3]

Beñat Oihartzabal [4]

Jasone Salaberria [4]

[1] Univ. del País Vasco/Euskal Herriko Unibertsitatea

[2] Elhuyar Foundation

[3] Gobierno Vasco/Eusko Jaurlaritza

[4] UMR 5478 IKER

Georg Rehm, Hans Uszkoreit
(editoreak, editors)

Editors

Georg Rehm
DFKI
Alt-Moabit 91c
Berlin 10559
Germany
e-mail: georg.rehm@dfki.de

Hans Uszkoreit
DFKI
Alt-Moabit 91c
Berlin 10559
Germany
e-mail: hans.uszkoreit@dfki.de

ISSN 2194-1416 ISSN 2194-1424 (electronic)
ISBN 978-3-642-30795-9 ISBN 978-3-642-30796-6 (eBook)
DOI 10.1007/978-3-642-30796-6
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012946621

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)



HITZAURREA

Liburu zuri hau hizkuntza-teknologiei eta haien potentzialei buruzko jakintza sustatzea helburu duen bildumaren atal bat da, hezitzaileei, kazetariei, politikariei eta hizkuntza-komunitateei zuzendua.

Europar, desberdina da, hizkuntza batetik bestera, hizkuntza-teknologiaren eskuragarritasuna eta erabilera. Horren ondorioz, desberdinak behar dute izan, halaber, hizkuntza bakoitzerako hizkuntza-teknologiaren ikerketa eta garapena bultzatzeko behar diren ekimenak.

Europako Batzordeak sortutako META-NET Bikaintasun Sareak gaur egungo hizkuntza-baliabideei eta teknologiei buruzko analisi bat bideratu du liburu zuriaren bilduma honetan (p. 75). Analisi hori Europako 23 hizkuntza ofizialentzako eta Europako beste zenbait nazio- eta eskualde-hizkuntza garrantzitsurentzako gauzatu da. Analisiaren ondorio gisa, ondorioztatu da ikerketa-hutsune esanguratsuak daudela hizkuntza bakoitzerako. Adituen gaur egungo egoeraren analisi eta ebaluazio xeheago batez, etorkizuneko ikerketen eragina handiagotu eta arriskuak gutxiagotu litezke.

Enpresa-munduko, administrazio publikoko, industria-sektoreko, ikerketa-alorreko, software-enpresetako, teknologia-hornitzaileetako eta unibertsitate europarretako parte-hartzaileekin lanean diharduten 33 herrialdetako 54 ikerketa-zentroz (p. 71) osatuta dago META-NET. Denak elkarrekin, teknologiari buruzko ikuspegi bateratu bat ari dira sortzen, eta, aldi berean, 2020 bitartean ikerketa-hutsuneak hizkuntza-teknologiaren bidez betetzeko bideak zein izan daitezkeen azaltzen duen ikerketa-agenda estrategiko bat ere ari da garatzen.

PREFACE

This white paper is part of a series that promotes knowledge about language technology and its potential. It addresses educators, journalists, politicians, language communities and others.

The availability and use of language technology in Europe varies between languages. Consequently, the actions that are required to further support research and development of language technologies also differ for each language. The required actions depend on many factors, such as the complexity of a given language and the size of its community.

META-NET, a Network of Excellence funded by the European Commission, has conducted an analysis of current language resources and technologies in this white paper series (p. 75). This analysis focussed on the 23 official European languages as well as other important national and regional languages in Europe. The results of this analysis suggest that there are many significant research gaps for each language. A more detailed expert analysis and assessment of the current situation will help maximise the impact of additional research and minimise any risks.

META-NET consists of 54 research centres from 33 countries (p. 71) that are working with stakeholders from commercial businesses, government agencies, industry, research organisations, software companies, technology providers and European universities. Together, they are creating a common technology vision while developing a strategic research agenda that shows how language technology applications can address any research gaps by 2020.

Dokumentu honen egileek beren eskerrik beroenak adierazi nahi dizkie alemanezko liburu zuriaren [1] egileei, haien dokumentuko zenbait atal, hizkuntzaren araberakoak ez direnak, berrerabiltzeko baimena emateagatik.

Liburu zuri hau Europako Batzordeko Zazpigarren Esparru Programaren eta IKTak Sustatzeko Programa Estrategikoa-ren diru-laguntzari esker garatu da, T4ME (249 119 Dirulaguntza Hitzarmena), CESAR (271 022 Dirulaguntza Hitzarmena), METANET4U (270 893 Dirulaguntza Hitzarmena) eta META-NORD (270 899 Dirulaguntza Hitzarmena) kontratuen baitan.

The authors of this document are grateful to the authors of the White Paper on German [1] for permission to re-use selected language-independent materials from their document.

The development of this white paper has been funded by the Seventh Framework Programme and the ICT Policy Support Programme of the European Commission under the contracts T4ME (Grant Agreement 249 119), CESAR (Grant Agreement 271 022), METANET4U (Grant Agreement 270 893) and META-NORD (Grant Agreement 270 899).



AURKIBIDEA CONTENTS

EUSKARA ARO DIGITALEAN

1	Laburpena	1
2	Arriskua, gure hizkuntzentzat, eta erronka, hizkuntza-teknologiarentzat	3
2.1	Hizkuntza-mugek oztopoak jartzen dizkiote Europako informazio-gizarteari	4
2.2	Gure hizkuntzak arriskuan	4
2.3	Hizkuntza-teknologia teknologia bideratzaile giltzarria da	5
2.4	Hizkuntza-teknologiaren aukerak	6
2.5	Hizkuntza-teknologiaren erronkak	6
2.6	Hizkuntzaren jabetzea gizakiengan eta makinetan	7
3	Euskara Europako informazio gizartean	9
3.1	Datu orokorrak	9
3.2	Euskararen berezitasunak	10
3.3	Azken gertaerak	11
3.4	Hizkuntza-lanketa	12
3.5	Hizkuntza hezkuntzan	12
3.6	Nazioartean	13
3.7	Euskara Interneten	13
4	Hizkuntza-teknologia euskararako	15
4.1	Hizkuntza-teknologia aplikatzeko arkitekturak	15
4.2	Aplikazio-eremu komunak	16
4.3	Beste erabilera-eremu batzuk	24
4.4	Hizkuntza-teknologia hezkuntzan	26
4.5	Hizkuntza-teknologiako programak	27
4.6	Euskararako tresna eta baliabideak	27
4.7	Hizkuntzarteko konparaketa	29
4.8	Ondorioak	30
5	META-NETi buruz	34

THE BASQUE LANGUAGE IN THE DIGITAL AGE

1	Executive Summary	35
2	Risk for Our Languages and a Challenge for Language Technology	37
2.1	Language Borders Hinder the European Information Society	38
2.2	Our Languages at Risk	38
2.3	Language Technology is a Key Enabling Technology	39
2.4	Opportunities for Language Technology	39
2.5	Challenges Facing Language Technology	40
2.6	Language Acquisition in Humans and Machines	40
3	Basque in the European Information Society	42
3.1	General Facts	42
3.2	Particularities of the Basque Language	43
3.3	Recent Developments	44
3.4	Language cultivation in Basque	45
3.5	Language in Education	45
3.6	International Aspects	46
3.7	Basque on the Internet	47
4	Language Technology Support for Basque	48
4.1	Application Architectures	48
4.2	Core Application Areas	49
4.3	Other Application Areas	56
4.4	Language Technology in Education	58
4.5	Language Technology Programs	59
4.6	Availability of Tools and Resources	59
4.7	Cross-language comparison	61
4.8	Conclusions	62
5	About META-NET	66
A	Aipamenak – References	67
B	META-NETeko Kideak – META-NET Members	71
C	META-NETen liburu zurien bilduma – The META-NET White Paper Series	75

LABURPENA

Hizkuntza gizakien arteko komunikazio-biderik garrantzitsuena da. Ideiak eta sentimenduak adierazteko aukera ematen digu, ikasten eta irakasten laguntzen digu, ezinbestekoa da bizitzeko, kulturaren transmisiorako tresnarik garrantzitsuena da, eta haren hiztunen identitate-ikurra da.

Hizkuntza gizakien arteko komunikazio-biderik garrantzitsuena da.

Gaur egun, mundu globalizatu honetan, edozein lekutako pertsonekin erraz komunikatzeko tresna asko ditugu. Adibidez, informazio- eta komunikazio-teknologia berriek sare sozialak garatzeko aukera eman dute, eta, hala, ekarpen handia izan da herrialde eta kultura desberdinetako pertsonak harremanetan jartzea sustatzeko. Azken urteotan, halaber, atzerritarren mugimendua handia izan da gure herrialdeetan, dela turismoagatik, dela immigrazioagatik, eta horrek hainbat hizkuntzatan komunikatzeko beharra sortu du. Harelere, hizkuntzarteko komunikazio-arazoak, maiz, lingua franca erabiliz gainditzen dira.

Europa aniztasun kultural eta linguistikoaren adibide garbia da, nahiz eta azken 60 urteetan zehar bateratzen politiko eta ekonomikoa izan duen. Hortaz, euskaratik polonierara zein italieratik islandierara, eragozpen linguistikoak gainditu behar dira, nahitaez, Europako hiritarren arteko eguneroko komunikazioan nahiz enpresaren eta politikaren esferetako komunikazioetan. Europar Batasunaren erakundeek bilioi bat euro behar dute urtean, beren eleaniztasun-politika betearazteko; alegia,

testuak itzultzeko eta ahozko jarduerak interpretatzeko. Bitartean, ingelesa lingua franca bihurtzen ari da Europako hiritarren arteko komunikazioan.

Espainiar estatuan ere antzeko eszenatokia dugu. Estatu osoan, hizkuntza ofizial bakarra dago: espainiera edo gaztelania; horiez gainera, hiru hizkuntza koofizial daude: euskara, galiziera eta katalana. Estatuan eleaniztasunari eustea ez da lan erraza izan; aitzitik, espainierak gainerako hizkuntzen artean duen gailentasunean, identitate kulturalaren eta linguistikoaren babesean oinarritutako prozesu konplexu baten emaitza izan da. Europako kasuan lingua franca gisa ingelesa erabiltzen den era berean, espainiera erabiltzen da maiz espainiar estatuko hizkuntza-eremu desberdinetako hiritarren arteko komunikazio zuzenerako.

Eleaniztasuna babestu beharreko kultura-ondarea da.

Bai Europari dagokionez, bai espainiar estatuari dagokionez, babestu beharreko kultura-ondarea da eleaniztasuna. Globalizazioak, lingua francaren erabilera gailendu eta gure hizkuntzaren erabilera murrizteko arriskua sortzen baitu, ez luke gure hizkuntza- eta kultura-ondare aberatsa alboratzen duen mekanismo bihurtu behar. Komunikazio-mundu global baten barnean, gure hizkuntza eta, harekin batera, gure identitate kulturala babesteko bideak aurkitu behar ditugu.

Gaur egungo hizkuntza-teknologiek eta ikerketa linguistikoek ekarpen handia egin dezakete eragozpen linguistiko horiek gainditzeko; izan ere, hizkuntza-